# 

## System Design

What are the system design considerations for your deployable ML model? Describe the iterations, delivery formats and limitations you may face and some solutions to overcome the limitations

* Should the model be deployed to run in batch, or to be hit from an api or some sort of streaming process as events are generated?
* What sort of infrastructure will be required for training? If it is a model that requires a lot of resources, where is the best place to train?

The data sets needed for the project will be used from below locations:

1. <https://www.aclweb.org/anthology/2020.nlposs-1.9.pdf>
2. <https://commonvoice.mozilla.org/en/datasets>
3. <https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html>

This project will be designed to detecting simple wake word for example “Ok Fourth Brain”. The project is confined to a single wake work detection in a single language only which is English. This involves two steps

1. The first step is to be able to convert an audio file to a text file using ASR – Automatic speech recognition algorithms. For now the hugging face ASR pipeline can be used to generate a text file from the audio file. This step involves audio model training.

**Hugging face ASR Pipeline:** [**https://huggingface.co/docs/transformers/v4.24.0/en/tasks/asr**](https://huggingface.co/docs/transformers/v4.24.0/en/tasks/asr)

1. The second step will be to read content from text file and compare it with defined wake work being considered here, example “Ok Fourth Brain”. This can be done by using language model training.

**Hugging face language model training:**

[**https://huggingface.co/blog/how-to-train**](https://huggingface.co/blog/how-to-train)

Diagram

Description automatically generated

The model would be run as a batch and it would not need too much resources as per the initial design.

## Ethical Considerations

Are there any ethical considerations of your project? Consider the data source, the intended outcome, and/or the eventual use cases.

* Did you modify anything about your plan based on these considerations?
* Can you anticipate any issues that might arise during the process?

This project does not seem to have any ethical considerations as the datasets are available publicly. I do not anticipate any issues as of now for this project as far as I know.